DOCUMENT RESUME

ED 111 258 HE 006 629

AUTHOR McLaughlin, Gerald W.
TITLE The Use of High School Faculty Ratings to Predict USMA Fourth Class Performance.
INSTITUTION Military Academy, West Point, N.Y. Office of Institutional Research.
REPORT NO IA4.06-71-001
PUB DATE Mar 71
NOTE 22p.

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS Academic Achievement; Achievement Rating; *Admission Criteria; Comparative Analysis; *Faculty Evaluation; Grade Point Average; *Higher Education; High School Graduates; *Military Schools; Predictive Measurement; Rating Scales; *Student Characteristics; Success Factors
IDENTIFIERS *United States Military Academy

ABSTRACT

The potential use of High School Faculty Ratings for admission purposes is investigated. The ratings include the evaluations of a candidate on 10 traits and three overall characteristics. The rating forms are given to a mathematics, English, and physical education teacher, a counselor or high school principal, and one other faculty member by the applicant, prior to his evaluation by USMA admissions officers. The sample of applicants studied included 697 candidates to the Class of 1972, 542 of whom were admitted. The 10 traits were combined into four for each of the five raters by factor analysis. Integer weights for raters were developed for each trait and then for the combined traits to best predict Fourth Class Aptitude for the Service and Fourth Class Grade Point Average. It was found that combinations of the ratings had significant validity for predicting both Fourth Class ASR and GPA. This was especially true with ASR for the group of cadets having ratings from at least a mathematics teacher, an English teacher, and a physical education teacher or coach. (Author/LBH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED111258

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to:
HE TM
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

UNITED STATES MILITARY ACADEMY

WEST POINT · NEW YORK

THE USE OF HIGH SCHOOL FACULTY RATINGS TO PREDICT USMA FOURTH CLASS PERFORMANCE

HE 006 629

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

DUTY-HONOR-COUNTRY

OFFICE OF INSTITUTIONAL RESEARCH

2

March 1971

THE USE OF HIGH SCHOOL FACULTY RATINGS TO PREDICT USMA FOURTH CLASS PERFORMANCE

IA4.06-71-001

by CPT Gerald W. McLaughlin

March 1971
Office of Institutional Research
United States Military Academy
West Point, New York 10996

ABSTRACT

This report investigates the potential use of High School Faculty Ratings for admission purposes. The ratings include the evaluations of a candidate on ten traits and three overall characteristics. The rating forms are given to a mathematics, English and physical education teacher, a counselor or high school principal, and one other faculty member by the applicant, prior to his evaluation by USMA Admissions Officers.

The sample of applicants investigated included 697 candidates to the Class of 1972, 542 of whom were admitted.

The ten traits were combined into four traits for each of the five raters by using factor analysis. Integer weights for raters were developed for each trait and then for the combined traits to best predict 4^{o} Aptitude for the Service and 4^{o} Grade Point Average. Several averaging techniques were investigated since all ratees did not have all raters, nor did all raters rate all traits.

It was found that combinations of the ratings had significant validity for predicting both 4^{o} ASR and 4^{o} GPA. This was especially true with ASR for the group of cadets having ratings from at least a mathematics teacher, an English teacher and a physical education teacher or coach.

PURPOSE

The purpose of this research is to investigate the validity of data from Faculty Ratings for predicting several aspects of 4° year academy performance. Its major emphasis is in obtaining a greater understanding of leadership performance at USMA as measured by the 4° ASR.

THE RATING FORM AND PROCEDURE

Prior to evaluation for admission, each candidate obtains several rating sheets to be given to faculty raters in his high school. These raters include a mathematics teacher, an English teacher, a coach or physical education teacher, a principal or counselor, and another faculty member ("Other").

a. Procedure. The raters rate the applicants on ten traits: (1) Seriousness of Purpose; (2) Responsibility and Dependability; (3) Moral and Ethical Values; (4) Industry and Application; (5) Cooperation and Teamwork; (6) Emotional Stability; (7) Common Sense and Judgment; (8) Bearing and Appearance; (9) Reaction to Criticism; and (10) Personal Magnetism. Ratings are also given on three overall recommendations: academic promise, character and personal promise, and overall recommendation. In addition to this information, the candidate's rank in class and the percent of the graduating class expecting to attend a four year college is obtained from a school official if possible.

b. Sample. Ratings on 697 candidates to the Class of 1972 were obtained from the Director of Admissions and Registrar. Of these 697 candidates, 128 were nominated but not admitted, 97 entered but failed to complete their 4° year, 55 more failed to complete the second year, and 417 finished at least two academic years. The sample was selected on a stratified random basis to assure representativeness of applicants in terms of acceptance and resignation, and includes about fifty percent of the members of the Class of 1972.

The following table shows the total number of raters of each type and the total number of applicants having a given type of rater.

TABLE 1

NUMBERS OF RATINGS & RATERS

	Number of Ratees Having Type of Rater	Total Number of Ratings by Type of Rater	Ratings/ Type of Rater	% of Ratees Having Type of Rater
MTH	569	739	1.30	81.6
ENG	561	688	1.22	80.5
PE/CO	377	413	1.10	59.3
*OTH	629	1104	1.75	90.2
**PRI/COUN	668	1282	1.91	95.8

METHODOLOGY

The basic methodological purpose of this research was to reduce the 65 possible scores for each candidate (5x13) to a smaller set of more reliable scores. It also sought to alleviate the problem of missing data. Both of these goals can be accomplished by combining homogeneous traits; that is, traits which are considered very similar by the raters. The average of several scores correctly weighted is always more reliable than any one of its components. Also, if a rater does not rate a specific trait, the rating can be estimated from the other scores in the group. The correlations used for condensing the traits within a type of rater were computed on the scores of a single rater per individual, rather than an average of all raters of that type for the ratees having at least one rater of the type. This was done since the more raters of a type per individual, the more reliable the average scores. The averaging thus produces a bias in the intercorrelations which is a function of the number of raters per ratee and which prevents a standardized comparison of trait intercorrelations across the raters.

The basic analysis was done by type of rater. The traits were correlated and factor analyzed for each of the five kinds of raters. The results were then used to graphically and statistically group the raw traits. The final groupings of traits were selected on the basis of the results within raters and also to assure consistency across raters. After the reduced set of

* This category includes 514 ratings on a total of 465 applicants from Science Teachers and 114 ratings from Military Science Teachers; categories which have been discontinued.

** Members of this class were requested to have both raters.

traits were formed, the score for a rater was the average of the scores he gave to the raw traits in the group. Next, raters of a given type were averaged for cases where an applicant had more than one rater of a given type. This produced a set of 35 scores for each applicant (5 raters each with 7 scores).

Once the raw traits are grouped, the analysis must deal with combining raters. The decision was made to group raters within trait. In other words, for a given trait, the raters were combined to predict the criteria.

The major advantage of grouping raters within traits lies in the problem of missing raters. Under this system, the missing score of a rater on a trait can be estimated from the other raters' scores on that trait. Thus, only a part of the final score will be estimated. On the other hard, if raters are used as the final variable and a rater is missing, then the entire variable score must be estimated, a less desirable occurrence.

The raters' scores on each trait were used to predict a criteria for each individual trait. The weights assigned to raters were integer weights which most consistently seemed to reflect the regression weights on the raters on a criteria for all of the traits. This was done because, in a sense, each trait is similar to every other trait with respect to the relative validities of the various raters. In other words, if the mathematics teacher is most valid on one trait, he will tend to be most valid on all traits, even though the validities of traits may vary. Under this type of situation, the variations in optimum weights of a rater over traits reflect sampling error. In light of this concept, the fact that the Class of '72 is a sample of all high school graduates in 1968, the fact that 1968 is itself a sample year, and from previous research in the profession using similar data, it was felt that integer weights provided the precision warranted at this point in the analysis.

Final trait scores were formed by summing the weighted scores of the raters on each trait. These final trait scores were then regressed against the respective criterion, thus obtaining the best prediction which could be obtained from a linear set of these predictors.

A final aspect of the research involved the investigation of two types of High School Rank. The first measure (HSR 1) was similar to the current measure expressed in a percent rather than a score ranging from 800 to 200. The second measure employs the same logic except that is uses the college bound segment of the graduating class as the effective class size rather than the entire graduating class. This was done for the following reason: HSR measures an individual's competitiveness with his peers within an academic system; therefore, the best estimate of the group with which he competes academically is the group planning to attend a four year college. This analysis does not differentiate between males and females, though this variable might also be relevant.

RESULTS

a. Grouping the Raw Traits. The intercorrelations between raw traits were factored for each type of rater. All raters and traits were not grouped together in a single analysis, since a correlation across raters excludes "halo" or rater bias error, while one within a given rater includes this bias or error. The average r across raters was about .2 to .3, while that within raters was about .6 to .8 showing the strength of the bias. This difference also reflects some systematic difference in the reliabilities of rater types, their agreement in trait definitions, and exposure to relevant ratee behavior.

The correlations between a rater's ratings on the ten raw traits were thus analyzed to determine if simplification and stability of the scores could be improved by grouping relatively homogeneous traits. The factor pattern was determined for each rater. These results are shown in the Appendix. Two dimensions were judged to adequately represent the correlations for each type of rater.

The first factor, defined by the raw trait ratings of "Seriousness of Purpose" and "Industry and Application" represents the ratee's propensity to work toward a goal and is named Perseverance. The second factor, defined by "Bearing and Appearance," "Personal Magnetism," "Emotional Stability," and "Reaction to Criticism" represents Leadership. The traits were then clustered visually and by using item intercorrelations corrected for the communalities of the items. This grouping produced four traits as shown in Table 2.

TABLE 2

COMPONENTS OF COMPONENT TRAITS

Number/New Trait	Number/Original Trait
1/Perseverance	1/Seriousness of Purpose 2/Responsibility & Dependability 4/Industry and Application
2/Situational Behavior	3/Moral & Ethical Values 5/Cooperation and Teamwork 7/Common Sense & Judgment
3/Charisma	8/Bearing & Appearance 10/Personal Magnetism
4/Receptiveness	6/Emotional Stability 9/Reaction to Criticism

The means of these and the overall ratings for each type of rater is shown in Table 3.

TABLE 3

TRAIT MEANS BY RATER

	Math	Eng	PE/Coach	Other	Prin/Coun	Average Standard Deviation
TRAIT 1	8.386	8.424	8.844	8.443	8.559	1.16
2	8.361	8.381	8.852	8.473	8.603	.99
3	7.997	8.021	8.381	8.129	8.180	1.16
4	8.132	8.270	8.498	8.297	8.345	1.23
Promise						
Academic	3.956	4.035	4.324	4.053	4.075	.76
Personal	4.287	4.348	4.562	4.350	4.440	.71
Overall	4.147	4.252	4.478	4.270	4.286	.73

(Traits scored on a 1-10 scale, overall on a 1 to 5 scale)

The main differences are that the PE/Coaches tend to give the highest ratings, then the Prin/Coun, English, and Mathematics Raters, in that order. Also, all raters tend to give higher ratings on the applied traits (1 & 2) and less so on the leadership type traits. Also, the means on the general ratings were highest for Personal Promise and lowest for Academic Promise. The overall ratings tend to be grouped at the higher end.

b. Internal Consistency of the Reduced Traits. To evaluate the internal consistency of the ratings, several criteria were applied (Campbell & Fiske, 1959):

(1) All intercorrelations between the same trait over different raters were significantly greater than zero, as were all but ten of the 420 correlations of different traits over different raters.

(2) For each trait, the correlations of the same trait-different raters should be higher than the different trait-different rater correlations for the corresponding raters. In other words, the correlation of Trait 1 rated by the mathematics rater and Trait 1 rated by the English rater should be larger than the correlations of Trait 1 (mathematics rater) and all other traits rated by the English rater. It should also be larger than the correlations of Trait 1 (English rater) and all other traits rated by the mathematics rater. In this case, there are ten correlations of the "same trait-different rater" type for each trait. This criterion is satisfied for

eight of the Trait 1 ratings, four of the Trait 2 ratings, five of the Trait 3 ratings, two of the Trait 4 ratings, seven of the Academic Promise ratings, two of the Personal Promise ratings, and one of the Overall ratings.

(3) The same trait-different rater correlations should be, but are not, higher than the different trait-same rater correlations. This indicates that a large part of each rating from a rater is caused by his overall impression of the ratee.

(4) The same pattern should exist among the different rater-different trait correlations as among the same rater-different trait correlations. Factoring of the different trait-same rater correlation showed that the rater had the same basic interpretation of the combined traits. Also, mean correlations were found for the different trait-same rater correlations and the correlations for different traits-different raters, as are shown in Tables 4 and 5. The diagonals in the different rater matrix (same trait-different rater) provide a lower bound for the inter-rater reliability of each trait.

The results show that the performance type ratings (in Trait 1 and Academic Promise) are more consistent than the personality type ratings (Personal Promise and Trait 4). A factor analysis was run, and the "same rater" and "different rater" correlation matrices had similar dimensions.

TABLE 4

AVERAGE CORRELATIONS FOR DIFFERENT TRAITS-SAME RATERS

	Trait 1	Trait 2	Trait 3	Trait 4	Overall 1	Overall 2
Trait 1	1.000					
Trait 2	.795	1.000				
Trait 3	.637	.730	1.000			
Trait 4	.702	.789	.710	1.000		
Overall 1	.642	.587	.496	.530	1.000	
Overall 2	.693	.717	.651	.670	.616	1.000
Overall 3	.702	.685	.603	.628	.753	.820

TABLE 5

AVERAGE CORRELATIONS FOR DIFFERENT RATERS

	Trait 1	Trait 2	Trait 3	Trait 4	Overall 1	Overall 2	Overall 3
Trait 1	.367						
Trait 2	.307	.317					
Trait 3	.254	.264	.310				
Trait 4	.275	.275	.248	.270			
Overall 1	.312	.263	.204	.217	.360		
Overall 2	.299	.290	.267	.269	.242	.294	
Overall 3	.297	.271	.232	.241	.279	.264	.284

c. Validity of Ratings. 4° ASR. Because of the range of correlations (validities) of the raters' scores with ASR within given traits, the raters' scores were regressed against ASR for each trait.

Various rules requiring a minimum number of raters were investigated. The results were compared with the situation where missing ratings were not estimated, and the resulting correlations were used as estimates of the population correlations. These techniques are discussed more fully in the Appendix.

The optimum decision rule balancing off the problems of representativeness and completeness of the data was to require that each ratee have at least three types of raters. On the average, each applicant was missing one rater, and this rater was usually the PE/Coach. The results of the analyses requiring at least three raters are shown in Table 6. Integer weights are used for previously discussed reasons. Academic Promise is not included, since it had no significant validity in predicting ASR.

TABLE 6

INTEGER WEIGHTS TO PREDICT 4° ASR
FOR THOSE WITH AT LEAST 3 RATERS

Rater	Trait 1	Trait 2	Trait 3	Trait 4	Personal Promise	Overall Recommendation
Math	2	2	1	1	3	3
English	-2	-2			-4	-3
PE/Coach	3	6	4	3	6	4
Other						
Prin/Coun						
Multiple R	.22	.26	.28	.21	.22	.19

The results show that Traits 2 and 3 had the highest validities, as indicated by Multiple R's of .26 and .28 respectively. These traits also had the largest potential validities if everyone were to have had a rater (missing data).

A simplified integer weighting scheme of (1, -1, 2) was then used for the first three raters for the seven traits. In addition, two scores were formed by unit weighting of the trait scores for the Prin/Coun and Other to see if these data might have validity in this form. The resulting trait validities and intercorrelations are given in the Appendix.

The resulting equation, where at least three types of raters were available, was: ASR' = 1.290 Trait 2 + 1.373 Trait 3 + 55.60 with a multiple correlation of .275 ($p < .05$). The overall scores from the Other and Prin/Coun failed to add any additional information. The correlations between the traits and ASR were about .02 lower than those from the correlations using regression weights from the trait regression. Thus, for this sample, the use of consistent integer weights is adequate, and the loss is less than the sampling error in the correlations which is about 0.04.*

10690

* $1/\sqrt{N-3}$ for small correlations and large N's.

However, these integer weights produce correlations as much as .15 lower than those obtained using the same weights in the missing data runs. To evaluate this slippage, the variables were further investigated. The major changes in means and correlations clearly showed that the coach was the rater most frequently missing from the rater set. This type of rater was the most valid in predicting ASR. To alleviate this confounding, the data were reanalyzed using only the 207 cadets having a Math, English, and PE/Coach Rater. This sample had more representative means on the traits and slightly lower standard deviations. The intercorrelations among the traits were about .03 lower than those in the missing data run, but the validities were representative. The results of the regression runs are shown in Table 7.

TABLE 7

INTEGER WEIGHTS TO PREDICT 4^{O} ASR FOR CADETS
HAVING MATH, ENGLISH & PE RATERS

Rater	Trait 1	Trait 2	Trait 3	Trait 4	Personal Promise	Overall Recommendation
Math	4	3	3		3	3
English	-3	3				
PE/Coach	4	6	5	3	7	5
Other						
Prin Coun						
Multiple R	.309	.318	.332	.230	.279	.222

The increase in the correlations are more in line with the validities expected from the missing data correlations.

The raters were again weighted (1, -1, 2) and the resulting variables were regressed against 4^{O} ASR. The resulting regression equation for this group was: ASR = 2.12 Trait 2 + 1.09 Trait 3 + 45.20 ($R = .337$, $p < .10$).

d. Validities of Ratings: 4^{O} GPA. The set of combined ratings were also regressed against 4^{O} GPA, using the same methodology as with ASR. In other words, all raters for a given trait were regressed against the criterion. The system requiring at least three raters was the only analysis performed. This was done because inspection of the missing data correlations showed that the Coach/PE rater was not uniquely valid for 4^{O} GPA, where he was for predicting 4^{O} ASR. The resulting weights are shown in Table 8.

TABLE 8

REGRESSION WEIGHTS FOR PREDICTING 4° GPA TRAIT

Rater	Trait 1	Trait 2	Trait 3	Trait 4	Academic Promise	Personal Promise	Overall Recommendation
Math	.026	.027	.015	.016	.047	.029	.047
English				.010			
Coach/PE					.034		.022
Other	.016	.016	.011	.013		.021	
Prin/Coun					.037		.019
Multiple R	.302	.245	.175	.216	.468	.205	.339

As would be expected, the most valid rater was the mathematics rater. This is most likely because of the heavy quantitative orientation of 4° academics. The "Other" category also tended to receive a larger weight. Much of this weight is due to the inclusion of the science raters in this category. Inspection of the zero order validities in the missing data analysis showed that the math rater was more valid, but that the other four raters were about equally valid on the average. On the basis of this evidence, trait scores were formed where the mathematics rater was given a weight of two and all other raters a weight of one. The resulting validities are shown in the Appendix. A comparison of the validities with the multiple correlations shows that the integer weights are adequate, at least for this sample. It should be noted that this fit is to 4° GPA which may be more quantitative than other grades.

The combined traits were then regressed against 4° GPA. The resulting equation was: $GPA' = .0270\ Ov\ 1 - .011\ Ov\ 2 + 2.0371$ ($R = .499$, $p < .01$).

This equation has a great deal of intuitive face validity, since the overall rating for character and personal promise acts to remove spurious bias from the overall rating for academic promise. This does not mean that the more disreputable a candidate, the higher the score. It merely shows that when a rater makes the judgment about an applicant's academic potential, the rater is also influenced or biased by the candidate's personality. Given two ratees with the same true academic potential, the one with the pleasing personality apparently receives the higher rating on academic potential.

e. Validity of the Ratings: Retention. The trait scores of the

raters were regressed against overall retention for the first two years at the Academy. Only the PE/Coach rater had consistent validity. The correlation was about .14 on the traits. Thus, this analysis was not continued because of the lack of statistical significance.

f. High School Rank. Two measures of a candidate's rank in his high school graduating class were investigated. The first measure was the applicant's percentile in his class. This rank is highly related to the current High School Rank, which has a mean of 500 and a range of 200 to 800. The second measure (HSR 2) is analogous, except it computes the candidate's rank in the college bound segment of his graduating class. In other words, the denomination includes only the number going to college, not the entire class. These two measures had the statistics shown in Table 9. Their intercorrelation was .775.

TABLE 9

STATISTICS FOR TWO MEASURES OF HIGH SCHOOL RANK

Correlations

	N	$\bar{X}$	SD	GPA	PE	ASR	PAE
HSR 1	695	82.7	16.2	.333	-.054	.029	-.075
HSR 2	291	69.4	23.3	.513	-.165	.007	-.214

Average Correlations with Traits by Rater

	Math	Eng	PE/Coach	Other	Prin/Coun
HSR 1	.281	.265	.145	.284	.349
HSR 2	.227	.228	.184	.293	.363

These results show that both measures are related to academics but relatively independent of the other academy measures. Also, they are about equally related to the ratings. The major differences are the correlations with 4° GPA and PAE, where HSR 2 has a stronger relationship with both measures. Unfortunately, too few cases with a measure of HSR were available to include it in other analyses.

FOOR

CONCLUSIONS

This research sought to determine the relative validities of high school raters on various criteria of Academy performance. The main criteria of interest were 4^{O} GPA and 4^{O} ASR. Retention was also included. It was found that various components of the ratings did have validity in predicting 4^{O} ASR and 4^{O} GPA.

First, the thirteen raw measures can be reduced to seven scores for each type of rater through the use of multivariate analysis. The raters are similar in their interpretation of the traits; however, increased definitions are required for clearer understanding of the concepts of Character, Personal Promise, and Overall Recommendation. "Receptiveness" (T-4), which includes "Emotional Stability" and "Reaction to Criticism," would also require additional clarification if it is to be retained; however, there seems to be no statistical reason for the retention of this measure.

In terms of predicting 4^{O} ASR, significant validity was obtained using the ratings of the math, English, and PE/Coach raters on the traits of "Situational Behavior" and "Charisma." This validity was significant when missing ratings were estimated from the available ratings ($r = .275$) but was higher when only applicants who had these three specific raters were included in the analysis ($r = .337$). The best equation will be found when "Charisma" is weighted one and "Situational Behavior" is given a weight of two, for the case where an individual has all three types of raters (math, English & Coach/PE.

The raters were also given weights to obtain trait scores for predicting 4^{O} GPA. In this case all raters were included on the traits. The use of a weighted combination of two times Academic Promise minus Character Promise was valid for predicting the criteria ($r = .499$). The ratings have no apparent validity for predicting either overall retention or physical performance.

A final analysis investigated a modified high school rank score. This measure, which gives rank in the college bound component of the graduating class, had about the same relationships with the ratings, but an appreciably higher relationship than the traditional measure with the criteria of 4^{O} GPA (.51 to .33).

RECOMMENDATIONS

The following recommendations are made:

a. Investigate the relationships between the measures derived from this research and other components of the pre-admissions data base.

b. Cross-validate the equations and weights from this study on the Class of 1973.

c. Require all future applicants to have ratings from a mathematics, English, PE/Coach and counselor/principal rater.

d. Delete the raw traits of "Emotional Stability" and "Reaction to Criticism" from the rating form, and develop additional measures of Situational Behavior and Charisma.

e. Clarify the concepts of Overall Recommendation, Character, and Personal Promise. Also, stretch out the upper end of the scale with selected adjectives to improve their internal consistency and statistical characteristics.

f. Encourage more data collection on the size of the college bound proportion of an applicant's graduating class to further investigate the validity of a HSR using this value in its computation.

REFERENCES

Campbell, D.T., and Fiske, D.W., "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix" Psycholcgical Bulletin, 1959, pp. 56, 81-105.

APPENDIX

1. Principal Axis Solutions of Raw Traits by Rater
2. Statistics for Traits After Weighting to Predict Criteria Measures
3. Sampling and Missing Data Procedures

10600

PRINCIPAL AXES SOLUTIONS OF RAW TRAITS BY RATER

MATH RATER

Trait	F_1	F_2	F_3	ho^2	h_2^2
1	.812	-.387	-.030	.810	.809
2	.821	-.208	-.189	.765	.717
3	.716	.014	-.059	.598	.513
4	.790	-.410	.068	.804	.792
5	.818	.121	-.145	.714	.689
6	.798	.178	.140	.704	.669
7	.762	-.038	.275	.667	.582
8	.719	.292	-.067	.692	.602
9	.817	.194	.039	.714	.705
10	.736	.297	-.025	.656	.630
Eigenvalue	6.083	.620	.168	7.074	6.703
% ho^2	86.0	94.8	97.1	100.0	94.8

ENGLISH RATER

Trait	F_1	F_2	F_3	ho^2	h_2^2
1	.824	-.316	-.110	.781	.779
2	.821	-.188	.083	.738	.709
3	.698	-.111	.311	.624	.500
4	.807	-.314	-.169	.781	.750
5	.829	.048	.010	.697	.690
6	.794	.134	.151	.697	.668
7	.763	.044	-.070	.627	.584
8	.688	.294	-.134	.635	.560
9	.807	.125	.028	.695	.667
10	.774	.278	-.091	.686	.676
Eigenvalue	6.114	.467	.196	6.961	6.583
% ho^2	87.8	94.5	97.4	100.0	94.6

PE/COACH RATER

Trait	F_1	F_2	F_3	ho^2	h_2^2
1	.801	-.260	-.075	.704	.709
2	.798	-.155	.157	.704	.661
3	.699	.003	.172	.556	.489
4	.698	-.297	-.024	.627	.575
5	.740	-.011	.046	.606	.548
6	.758	.177	.064	.625	.606
7	.786	-.003	-.140	.640	.618
8	.671	.188	-.211	.552	.486
9	.691	.266	.192	.592	.548
10	.731	.140	-.181	.596	.554
Eigenvalue	5.457	.337	.200	6.202	5.790
% ho^2	88.0	93.4	96.6	100.0	93.4

OTHER RATER

Trait	F_1	F_2	F_3	ho^2	h_2^2
1	.853	-.330	-.051	.830	.837
2	.862	-.139	.045	.793	.762
3	.739	.076	.289	.654	.552
4	.829	-.364	-.023	.830	.820
5	.816	.031	.200	.719	.667
6	.850	.226	-.024	.750	.774
7	.818	.087	-.232	.715	.677
8	.763	.108	-.160	.675	.594
9	.838	.133	-.011	.750	.720
10	.829	.198	-.014	.736	.726
Eigenvalue	6.736	.395	.208	7.452	7.129
% ho^2	90.4	95.7	98.5	100.0	95.7

SCIENCE RATER

Trait	F_1	F_2	F_3	ho^2	h_2^2
1	.815	-.335	-.012	.764	.776
2	.783	-.251	.184	.729	.676
3	.679	.030	.286	.570	.462
4	.794	-.323	-.086	.764	.735
5	.801	.069	-.167	.693	.646
6	.782	.219	.094	.678	.660
7	.748	-.028	-.148	.624	.560
8	.726	.266	.007	.644	.598
9	.788	.203	.098	.678	.662
10	.786	.187	-.176	.693	.653
Eigenvalue	5.947	.481	.213	6.837	6.428
% ho^2	87.0	94.0	97.1	100.0	94.0

PRIN/COUN

Trait	F_1	F_2	F_3	ho^2	h_2^2
1	.804	-.349	.080	.766	.768
2	.846	-.179	.094	.766	.748
3	.696	-.021	.334	.616	.485
4	.749	-.365	.103	.729	.694
5	.819	.075	.065	.691	.676
6	.820	.233	.073	.707	.727
7	.782	.071	-.065	.644	.617
8	.706	.185	-.131	.604	.533
9	.816	.167	-.018	.716	.694
10	.762	.191	-.165	.660	.617
Eigenvalue	6.108	.450	.196	6.899	6.559
% ho^2	88.5	95.1	97.9	100.0	95.1

ho^2 = estimated communality, largest correlation; h_2^2 = communality of first two factors; Eigenvalue = sum of squared loadings on a factor, the amount of variance accounted for by a factor; % ho^2 = the cumulative variance accounted for by a factor & all preceding factors; F_1, F_2, F_3 = the first three principal axes; Eigenvalue of ho^2 = the trace, the total amount of common variance in a set of correlations.

STATISTICS FOR FACULTY RATINGS WEIGHTED TO PREDICT
4° GPA (2,1,1,1,1)* 3 OR MORE RATERS

Correlations (N = 442)

Trait	Mean	S.D.	4°GPA	TR1	TR2	TR3	TR4	O_V1	O_V2	O_V3
1	51.49	4.82	.301	1.000						
2	51.44	3.85	.247	.823	1.000					
3	49.18	4.50	.176	.660	.764	1.000				
4	50.24	4.72	.200	.758	.847	.759	1.000			
O_V1	24.84	3.13	.470	.780	.700	.563	.617	1.000		
O_V2	26.58	2.76	.198	.780	.810	.740	.775	.679	1.000	
O_V3	25.97	2.87	.336	.799	.768	.663	.697	.824	.846	1.000

STATISTICS FOR FACULTY RATINGS WEIGHTED TO PREDICT
4° ASR (1,-1,2,0,0)** 3 OR MORE RATERS

Correlations (N = 491)

Trait	Mean	S.D.	4°ASR	TR1	TR2	TR3	TR4	O_V1	O_V2	O_V3
1	17.56	2.20	.181	1.000						
2	17.54	1.86	.248	.745	1.000					
3	16.69	2.32	.260	.628	.710	1.000				
4	16.89	2.04	.207	.665	.722	.672	1.000			
O_V1	8.92	1.49	.116	.570	.513	.425	.479	1.000		
O_V2	8.99	1.41	.202	.630	.682	.628	.618	.573	1.000	
O_V3	8.76	1.36	.155	.617	.589	.520	.528	.681	.765	1.000

STATISTICS FOR FACULTY RATINGS WEIGHTED TO PREDICT
4° ASR (1,-1,2,0,0); MATH, ENGLISH & PE/COACH RATERS

Correlations (N = 207)

Trait	Mean	S.D.	4°ASR	TR1	TR2	TR3	TR4	O_V1	O_V2	O_V3
1	17.91	2.36	.300	1.000						
2	18.03	1.98	.318	.724	1.000					
3	17.13	2.58	.291	.620	.642	1.000				
4	17.17	2.76	.245	.649	.681	.640	1.000			
O_V1	8.72	1.62	.142	.465	.400	.362	.431	1.000		
O_V2	9.31	1.52	.289	.567	.653	.599	.608	.509	1.000	
O_V3	9.00	1.44	.225	.521	.517	.457	.483	.590	.705	1.000

* Math Rater weighted two, all others unit weights.

** Math Rater weighted one, English weighted minus one, PE/Coach weighted two, Other & PRIN/COUN weighted zero.

SAMPLING AND MISSING DATA PROCEDURES

One of the problems with any system utilizing ratings as predictors is missing data. Some raters do not rate all traits, nor do all ratees have ratings from all raters. The first problem can be alleviated by combining raw traits and using the available raw trait scores to estimate the scores on other traits in that group which have not been rated.

The second problem is more serious, since frequently there is little information on which to base an estimate. Use of regression analyses to estimate any score from available raters is one solution. However if people were required to have at least four of the five raters, this would require five equations for each trait or 35 equations. If only three of the five raters were required, 15 equations would be needed to account for the different ways a ratee could have three or more of five raters. For seven traits, this would require 105 equations. Thus, while this is one of the more accurate methods, it is not feasible in this case.

Another method is to estimate the missing scores as the mean score given by the type of rater who is missing on each missing trait. This method is a very gross estimate and thus desirable only in terms of simplicity.

A third method lies between these two extremes and uses the available scores from other raters on a trait to estimate the missing ratings on that trait. A variation uses the available standard scores on the trait to estimate the missing standard score which is then translated into the missing raw scores. This variation is desirable where there are large variations in the means and standard deviations of the raters' scores within a specific trait. This was not the case here, with the possible exception of the PE/coach rater who consistently has means about 1/3 of a standard deviation above the other raters' means.

In light of these conditions, this method of estimation was used in estimating missing data from the raw scores. As would be expected, the mean of the PE/coach ratings went down and the standard deviation increased slightly, since this was the most frequently missing rater. These, however, were not serious changes.

The estimation of missing data did cause the intercorrelation among the raters to increase rather markedly, and also caused the validities of the raters to decrease. To determine the feasibility of various administrative requirements of the number of raters, runs were made using all candidates with all five raters, with at least four raters, and at least three raters. The results were compared with the intercorrelations and validities of the measures where no estimations were made. In other words, the correlations between measures were computed on all individuals who had both measures. This missing data analysis provides a type of optimum result that would have been obtained if the presence of pairs of measures was a random sample from the population.

In the analysis of 4° ASR, it was found that requiring three or more raters provided validities most similar to those in the missing data. The use of "at least four" requirement produced a lower set of validities with no accompanying reduction of correlations between the predictors. If this reduction had occurred, the final multiple correlation would have remained stable. Even though the individual validities decrease, their uniqueness would increase. The final weights for raters, however, were very similar to the system requiring "at least three." The requirement that the applicants have all scores produced a sample too small for consideration (N=120) with erratic statistics.

Thus, even though the validities were lower for the "at least three" than for the "missing data" analysis, it was the best of the three decision rules. It was further evaluated by using only those with the raters given weights to predict 4° ASR. These weights from the "at least three" produced validities very similar to those from the missing data.

In the future, all applicants should be strongly encouraged to obtain a rating from all rater types; but if data are missing, there will be reason to feel that at least three raters will produce scores with acceptable validity.